DOCUMENT RESUME

ED 320 350 EC 231 159

AUTHOR Kennedy, Mary F.; And Others
TITLE On Doing a Naturalistic Evaluation.
PUB DATE 85
NOTE 19p.
PUB TYPE Reports - Descriptive (141) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Enrichment Activities; *Evaluation Methods; Foreign Countries; *Gifted; *Naturalistic Observation; *Program Evaluation; *Qualitative Research; *Research Methodology
IDENTIFIERS Newfoundland

ABSTRACT

The paper describes a naturalistic evaluation study of an enrichment program for gifted learners in Newfoundland. The pull-out program (for two half days out of six) is described in terms of audience concerns and issues, standards, program components, and data-generating techniques (observations and interviews). The evaluation led to a report pointing to specific strengths and weaknesses and went beyond assessment of the impact of the program to thorough description and judgment. Also discussed is the naturalistic approach to program evaluation in general considering its emergent design, rigor factors inhibiting naturalistic evaluation (such as the non-quantitative results), cost factors inhibiting naturalistic evaluation, and training factors inhibiting naturalistic evaluation. Contains 10 references. (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED320350

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Mary Kennedy

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

On Doing a Naturalistic Evaluation

by

Mary F. Kennedy
Lily K. Abbass
June P. Perry

Introduction

From theory to practice is frequently a risky venture, and it is especially the case when the theory is, to any extent, in confrontational stance with traditional theory as held by the majority of practitioners in a given field. In considering approaches to evaluation, the authors have been, in theory and in practice, adherents of the naturalistic approach as delineated by Guba and Lincoln (1981). During the past five years the authors have completed four evaluation studies, some separately and some together. It is the latest study, the evaluation of an enrichment program for gifted learners, which will provide the framework for the discussion of naturalistic evaluation in practice.

The Program

In September, 1983 the Roman Catholic School Board for St. John's implemented an enrichment program, on a pilot basis, for academically gifted learners. The school board is large and urban, with a student population of approximately 20,000 attending 36 schools. Despite the religious affiliation the school board is public; in the province of Newfoundland the government funds all education through the various religious denominations.

The enrichment program implemented by the school board was the first in the province of Newfoundland. The school board adopted the definition of giftedness proposed by the United States Office of Education:

EC 231157

BEST COPY AVAILABLE

> Gifted and talented children are those identified by professionally qualified persons who by virtue of outstanding abilities, are capable of high performance... Children capable of high performance include those with demonstrated achievement and/or potential ability in any of the following areas, singly or in combination: (1) general intellectual ability; (2) specific academic aptitude; (3) creative or productive thinking; (4) leadership ability; (5) visual and performing arts; (6) psychomotor ability
>
> (Marland, 1972)

For the pilot period it was decided to focus on the intellectually and academically gifted students only, since government at that time did not have policy on gifted education, nor did it fund enrichment programs. The school board was unwilling to wait for government to formulate policy, yet it realized that resources would be severely limited until government took fiscal responsibility for the program.

The enrichment program began on a small scale. Students from four schools were tested at the grade five level. Testing, identification, and selection procedures were completed by January, 1984, and the first group of twelve students began attending enrichment classes in February, 1984. During the next year four additional schools were tested, and two classes of approximately twelve students each were in operation, comprised of students at the grade five and six level. It was at this point that the school board requested evaluation assistance.

No financial allocation had been made for evaluation of the program. The senior author, who was a professional colleague of administrative personnel at the school board, was approached for assistance because of her known research interest in evaluation. The author immediately complied, in that the enrichment program provided an opportunity to do

a large-scale program evaluation. Research funds were obtained from Memorial University of Newfoundland, the place of employment of the senior author, and two graduate students, both experienced in evaluation, joined the evaluation team.

The Setting

The enrichment program, at the pilot stage, followed a withdrawal enrichment design. Students attended regular classes, and were withdrawn for two half days of a six day cycle; during withdrawal periods they attended enrichment classes. Enrichment classes were located at a linked complex, St. Pius X Elementary and Junior High Schools. The complex was selected as the original site for enrichment classes on the basis of its central location and the interest expressed by school administrators. Both schools provided classroom space and access to learning centre facilities.

At the pilot stage no attempt was made to include enrichment activities within the regular classes, nor were students other than academically gifted provided with enrichment activities, although long-term plans encompass both types of activities. The program was restricted, at the pilot stage, by two factors: (1) the lack of available funding for enrichment programs at the provincial level required that all costs must be gleaned from existing operating grants, and (2) the school board decided to proceed with caution in order to build a worthwhile program.

The goals of the enrichment program, as gleaned from program documents, are as follows:

1. To provide a differentiated curriculum which offers stimulating and challenging learning opportunities to children identified as being gifted.
2. To develop skills in productive thinking.
3. To develop skills in creative thinking.
4. To develop skills in creative problem-solving.
5. To develop research skills, allowing independent learning and investigation of real problems.
6. To explore topics in mathematics beyond the regular curriculum.
7. To develop the ability to express ideas clearly and fluently both orally and in writing.
8. To promote growth of a positive self-concept and interpersonal skills.
9. To develop the ability to listen well and work with others.
10. To develop skills enabling student to examine cultural and societal values.
11. To develop the ability to make sound moral judgements.

(Interim Report, 1985)

The program was staffed by a program coordinator at the district office level, and two enrichment class teachers. Both teachers dealt with four different groups of learners, and they taught four days of a six day cycle. The two non-teaching days were used for administration, planning, and liaison purposes.

The Evaluation

The evaluation team began its preliminary work toward the end of the two year pilot period, and continued into the current school year. The evaluation was emergent in design, as naturalistic, responsive evaluations should be, with data generated in each activity dictating the next set of evaluation activities. Evaluators worked on the project on a part-time basis only, having, for the most part, full-time work commitments elsewhere. In all, seven months elapsed from preliminary reading and analysis of documents to the preparation of final reports.

Audience Concerns and Issues

The evaluation was to be guided by the information needs of various audiences, as suggested by Stake (1967). Audiences are "those persons entitled by virtue of holding a stake to propose concerns and issues and to receive a report responsive to them" (Guba and Lincoln, 1981, p.37). Eight audience groups were identified. They included school board personnel responsible for the program; parents; students; regular classroom teachers whose students attended enrichment classes; school principals; enrichment program teachers; executive members of the provincial association for the gifted; representatives of the special education division of the provincial department of education.

Concerns and issues were elicited from these audiences. Approximately half of each group received a brief questionnaire seeking clarification of their role in the program, their concerns, issues, and expectations, and their opinion as to the main indicator of effectiveness in judging the program. Questionnaire recipients were then interviewed by telephone, when possible, or asked to return the completed questionnaire. Semantic

content analysis (Krippendorf, 1980) was used to analyse data on concerns and issues.

Standards

Guba and Lincoln (1981) state "judgments cannot be rendered in the absence of standards" (p.353). In traditional approaches to evaluation program objectives function as a clearly stated set of standards. In the naturalistic approach it is neither desirable nor feasible to use a set of preformulated objectives as the sole criteria in judging the entity being evaluated. The naturalistic evaluator must derive standards.

In evaluating the enrichment program criteria were established from a variety of sources. Guba and Lincoln (1981) suggest as one possible source "substantive experts" (p.353) Those knowledgeable in the area of gifted education, both in person and through the literature, provided appropriate criteria. Examinations of program goals provided another source from which to formulate criteria. And the audiences, through their concerns and issues, provided criterion statements. All data were incorporated in a comprehensive set of standards for all components of the enrichment program.

Program Components

> A comprehensive evaluation should broaden the scope of issues raised and seek answers to questions which related to the overall functioning of the program. In fact, it should be so broadened as to include not only assessments of the impact of the program on its clientele, i.e., gifted students, but also attempt a thorough description of the program's actual components.
>
> (Callahan, 1982, p.3)

In planning evaluation activities, it became evident early on that the complexity of the program would lead to chaos in collecting and

analysing data unless the program was divided into manageable units. Through analysis of documents and records eleven distinct program components were identified. They include (1) program goals and objectives; (2) identification and selection procedures; (3) curriculum; (4) classroom activities; (5) facilities and resources; (6) staff; (7) students; (8) evaluation procedures; (9) reporting procedures; (10) inservice training; (11) communication procedures. Standards and potential data sources were delineated for each component of the program.

Data-Generating Techniques

Observations. Observations were carried out over a two month period in time. Site visitations were done by two members of the evaluation team; each visitation was of one half day duration. The first five observation periods were open or non-directed. Evaluators made use of these visits to grow accustomed to the setting, to permit the students and the enrichment teacher to grow accustomed to their presence, and to determine what types of things to look for in the future.

Each observation period was followed by an interview with the enrichment teacher. Notes taken during the observations were read to the teacher for confirmation of evaluators' perceptions. The interviews also permitted the enrichment teacher to voice her opinions and concerns; these often pinpointed specific aspects of the classroom life to observe in the future.

Following the first set of open observations, classroom visitation periods became more structured in that evaluators sought specific data-observations of the communications flow within the classroom setting across a variety of activities; observations of student time on task;

observations of group interactions; observations of environmental factors pertaining to openness and group composition. These observations comprised an additional six half-day visits, in which both evaluators worked separately. Again data were shared with the enrichment teacher following each observation period. The observations permitted thick descriptions of classroom occurrences on a daily basis.

Interviews. In addition to a number of elite interviews with the enrichment teachers and the program coordinator, interviewing was used extensively throughout the evaluation. Interviews ranged from elite to semi-structured. The latter were conducted by following a comprehensive guide, yet permitting both the interviewer and the subject to redirect the order or to add items as required. When highly structured information was sought, written questionnaires were used.

In all more than half of the students in the enrichment pilot were interviewed; their opinions were sought on matters ranging from their perception of peer reaction to their selection, to program objectives, classroom activities, program resources, evaluation procedures, and their feelings about being identified and selected for the program. Students not interviewed completed a questionnaire seeking their opinions on the enrichment program. All students also completed a classroom activity checklist, designed to permit comparison of enrichment class and regular class activities.

Interviews were also conducted with others; these included specialists within the provincial department of education, a representative from the provincial association on giftedness, and pertinent school board

personnel. These interviews permitted free and open discussion of issues surrounding the program.

Information gleaned from analysis of program documents and records, observations, and interviews guided the development of questionnaires. Questionnaires were utilized with all members of the following audiences: regular classroom teachers, school principals, enrichment program personnel, and parents. The instruments varied in accordance with information sought from each audience, but all had certain elements in common. These include a Likert Scale to survey attitudes, a checklist on specific elements of the program, and a section allowing for open-ended responses.

Summary. Data-generating techniques were utilized to elicit new information, but also to confirm information gleaned from other sources. The data generated throughout the evaluation were plentiful and rich. Through the data a complete picture of the program emerged, enabling evaluators to address the concerns and issues of all audiences, and to delineate specific areas of strength and weakness so that improvements can be implemented.

The Naturalistic Approach

A naturalistic evaluation is characterized by an emergent design, responsiveness to audiences, and reliance on sociological research methods yielding, for the most part, qualitative data. Such an evaluation is discovery-oriented, resulting in thorough description, as well as judgement, of the entity being evaluated. With this definition in mind, it is apparent that the evaluation of the enrichment program

summarized in this paper can indeed be categorized as naturalistic. It is of interest, then, to consider the approach in terms of its application, its fit, to program evaluation.

Rigor Factors Inhibiting Naturalistic Evaluation

In deciding on a naturalistic approach the first problem one is likely to encounter is the non-quantitative results, which are viewed by many clients or audiences as non-scientific, hence non-credible. It is a fact that evaluators work in a milieu where great faith is invested in "scientificness" and any attempt to veer from the scientific, quantitative approach is viewed with askance.

In this particular circumstance the client presented no problem in deciding on the type of approach, since the evaluators were not paid by the client, and in fact had themselves generated other funds for the evaluation. Nevertheless the evaluators were aware that various audience information needs were to be met, and these audiences were likely to hold the same scientific bias that is evident across education and/or research groups. Hence care was taken, at all stages of the evaluation, to ensure rigor, in an effort to enhance the credibility of the results.

Guba (1978) suggests seeking to establish credibility, as opposed to the internal validity sought by the scientific paradigm. In evaluating the enrichment program credibility was established through prolonged interaction with the evaluation setting on the part of two evaluators. These interactions took place over a two month period on a regular basis. Each site visit was followed by a group debriefing session. At these sessions the third team member, who was not at the site, was invaluable in demanding clarifications, explanations, and in focussing

on specific aspects in need of further investigation. Following observations and interviews, all notes were summarized and interpretations derived; these were then checked with data sources for correctness.

Eisner (1979) defines structural corroboration as "the process of gathering data or information and using it to establish links that eventually create a whole that is supported by the bits of evidence that constitute it" (p.215). Structural corroboration can be established through triangulation; the use of many data sources and data-generating techniques, parts of which might be designed specifically to provide corroborative evidence, so that information can be consistently confirmed by other sources. In the enrichment program evaluation, data gleaned through the analysis of documents and records and observations were confirmed through interviews; they in turn were checked through questionnaires across groups so that, at the conclusion, multiple sources had reiterated much of the same data.

Guba and Lincoln (1981) note that the concept of generalization is often meaningless in evaluation, since evaluations are site-specific. However, this is not the case in the enrichment program evaluation. It is true that the purpose of the evaluation was to establish the merit and worth of the program itself, but it is the first enrichment program in the province of Newfoundland, and many groups are very interested in the results. These include school boards seeking to establish such programs for gifted students, and the provincial department of education. The latter is in the process of formulating policy on the education of the gifted. Hence reports could be used as leverage to seek funding from governing bodies in the future.

Guba and Lincoln (1981) suggest that generalization be replaced by fittingness; the degree of fit among evaluation contexts. The naturalistic approach allows for the heavy description needed to establish fittingness. In the case of the enrichment program evaluation, enough data were generated to permit description of all program components from the varying perspectives, attitudes, and values of all audiences.

While consistency, or replicability, cannot be readily established in naturalistic evaluations, step-wise replication can be undertaken (Guba and Lincoln, 1981). Throughout the evaluation of the enrichment program this procedure was used. Each stage of the evaluation, as it emerged, was planned so that data sources were divided among team members, who then worked independently. Observations and interviews were always completed by two team members functioning independent of each other. Documents and records were analysed and questionnaire pools generated using the same approach. All three team members worked separately in the initial stage; the team would then pool information for consistency cross-checks. In this manner team members, working independently, corroborated one another's work.

Summary. Attempts made to ensure rigor in the naturalistic approach to evaluation of the enrichment program might fail to convince the true disciple of the scientific paradigm that such evaluations meet the demands of validity, reliability, and objectivity. However, they do ensure that the evaluation will provide complete and comprehensive description of the program, and judgements that are based on sound knowledge of the program and its functions.

Cost Factors Inhibiting Naturalistic Evaluation

Guba and Lincoln (1981) state "sponsors want evaluation designs to be specified in great detail in order to reassure themselves that the study will be cost efficient and rigorous" (p.275). In the case of the enrichment program evaluation the client was not concerned with problems of cost, since the evaluation was conducted at the expense of the evaluators' institution, in support of research interests. Nevertheless the financial cost of evaluations is indeed an important factor in determining which types of evaluation are performed; little program development and implementation costs are earmarked for evaluation in most circumstances.

It has been estimated that the evaluation of the enrichment program, completed over a seven month period, comprised actually six weeks of full-time work for the evaluation team. In terms of expected earnings at current salary levels, as opposed to much higher consultant fees, this represents approximately $11,000. In addition $600. administrative costs were incurred, including typing, duplicating, and mailing. In implementing a naturalistic evaluation the cost, then, is incurred in personnel; the evaluators, if they are to function within the evaluation setting for a prolonged period of time, become very costly data-generating instruments. The point must be made, in the case of the enrichment program evaluation, that the evaluators had few, if any, real financial constraints, hence they were able to do exactly as they desired. The evaluation, using a naturalistic approach, could probably have been completed for significantly less financial cost, with the same results. But rigor, established through structural corroboration, triangulation,

overlap methods and stepwise replication, would also have been significantly less.

While a similar costing has not been estimated for a more traditional approach - one requiring the administration of paper and pencil instruments and little environmental interaction, it is probable that the cost would be diminished by one-half to two-thirds. It is certainly safe to predict that the total cost of a more conventional approach would be under $6000. So clients, obviously, must be convinced of the increased utility and value of the results of naturalistic evaluations if they are to be required to absorb significantly greater financial cost.

Training Factors Inhibiting Naturalistic Evaluation

Stake and Denny (1969), in considering the need for new approaches to evaluation, stated:

> We wish to acknowledge that the successful use of methods depends on the people who use them. We will point out that the specific technical skills suited to educational testing and educational research are not (to the surprise of some readers) perfectly suited to educational evaluation.
>
> (p. 370)

Even today the main training ground for those who function as evaluators is educational research/measurement. Competencies required by the naturalistic evaluator, beyond those required by the educational researcher, include observation and interview skills, openness, listening skills, the ability to work from tacit knowledge, on the spot information-processing skills, non-verbal communication skills; a whole set of competencies that might be classified as responsive, empathetic behaviour. Guba and Lincoln (1981) state:

> But if evaluators do not measure, what, then, do they do? They do what anthropologists, social scientists, con-

> noisseurs, critics, oral historians, novelists, essayists, and poets throughout the years have done. They empathize, describe, judge, compare, portray, evoke images, and create, for the reader and listener, a sense of having been there.
>
> (p.149)

It is therefore essential that the naturalistic evaluator learn to attend to the subjective, to get in touch with his own biases, and to trust his own judgement.

Other skills and competencies involve the analysis of qualitative data, which is neither as direct nor as easy as analysis of quantitative data. The analysis of qualitative data demands much greater interpretive skill on the part of the evaluator. Some form of content analysis is required, and it is only through practice with a given categorization scheme that the evaluator becomes facile. Throughout the enrichment program evaluation the authors relied on semantic content analysis as delineated by Krippendorf (1980); it proved to be the key to manageable data, without which chaos would have resulted.

Instrument design has long been the domain of the educational researcher, and there is no doubt that those with educational measurement and research backgrounds are experienced in the development of questionnaires as well as the reliability-testing of such tools. But it is not quite the same task to develop open-ended questionnaires, especially those which require lengthy interaction on the part of respondents. Such open-ended instruments should incorporate principles of print layout and design, so that respondents can complete them with as much ease as possible. The content of questionnaires, in the naturalistic approach, is also liable to differ considerably from questionnaires normally used in the scientific approach, and might incorporate oase

studies and scenarios in order to elicit information indirectly. These techniques are especially valuable in generating attitudinal and values data. The naturalistic evaluation might at least, then, require significantly different skills in instrument design than those usually developed working within the scientific approach.

Conclusions

Callahan (1982) summarizes the frustrations experienced because of the traditional approach to evaluation:

> How often an evaluator has concluded that a program 'has been effective' or 'doesn't bring about student change' without having any idea of what the program actually was, what factors impacted upon its success or lack of success, or of ways it might be improved.
>
> (p.3)

Doing a naturalistic evaluation is not straight-forward, easy, quick, or inexpensive. It can be thought of, in fact, in polar terms: convoluted, difficult, slow, and expensive. It takes time to immerse oneself in the setting, to study all aspects of a program, to discover nuances, to deal on a one-to-one basis with the human beings involved, to answer the information wants and needs of diverse audiences. But the pay-off is worthwhile.

The naturalistic evaluation of the enrichment program, over a seven month period, yielded immense and rich data. All components of the program were evaluated, and the client report pointed to specific strengths and weakness, so that improvements could be made. The evaluation looked beyond assessment of the impact of the program to thorough description and judgement. The naturalistic approach facilitated the emergence of the unique facets of the gifted program "a mixture of

people, politics, history, context, resources, constraints, values, needs, and interests" (Patton, 1981, p.23).

Selected References

Callahan, C.M. (1982). Issues in evaluating programs for the gifted. In Gifted Child Quarterly, 26 (2), 3 - 7.

Eisner, E.W. (1979). The educational imagination. New York: MacMillan.

Guba, E.G. (1978). Toward a methodology of naturalistic inquiry in educational evaluation. Monograph Series #8. Los Angeles: Center for the Study of Evaluation, University of California.

Guba, E.G. and Lincoln, Y.S. (1981). Effective evaluation. San Francisco: Jossey-Bass.

Interim Report. (1985). Enrichment program. Roman Catholic School Board for St. John's.

Krippendorf, K. (1980). Content analysis. Beverly Hills: Sage.

Marland, S. (1972). Education of the gifted and talented. Report to the Congress of the United States Commissioner of Education. Washington, D.C.: United States Government Printing Office.

Patton, M.Q. (1981). Creative evaluation. Beverly Hills: Sage.

Stake, R.E. (1967). The countenance of educational evaluation. In Teachers College Record, 68, 523-540.

Stake, R.E. and Denny, T. (1969). Needed concepts and techniques for utilizing more fully the potential of evaluation. In R.W. Tyler (Ed.), Educational evaluation: new roles, new means. Chicago: National Society for the Study of Education.